Aerosol mass spectrometer constraint on the global secondary organic aerosol budget
D. V. Spracklen et al.

We thank the reviewers for their comments on our paper. To guide the review process we have copied the reviewer comments in italics. Our responses are in regular font. We have responded to all the referee comments and made alterations to our paper (in bold text).

Anonymous Referee #1
Overview
This work uses measurements from AMS instruments, IMPROVE and 14C datasets, along with a global aerosol model, to constrain the total SOA budget as well as attribute the observed SOA to precursors from biogenic vs anthropogenically controlled sources. The paper is very clear and well written, and presents detailed sensitivity and uncertainty analysis of the model calculations. The authors argue that there is a substantial amount of SOA, 100 Tg / yr, which is governed by anthropogenic processes, and from 14C analysis estimate that a bulk of this is biogenic VOCs whose SOA yields are anthropogenically induced. As well, the implications of the SOA budget are extrapolated to impacts on aerosol radiative forcing, which, while an interesting estimate, feels a bit superfluous to the main objectives of SOA source attribution. Overall, the approach presented is new and timely, and a good fit for ACP. There have been a few other recent works which have attempted to place bounds on SOA budgets, and this work compliments these efforts and takes a step in a new and valuable direction. Still, I have several concerns about the methods, which seem a bit ad hoc, and the interpretation of the results, particularly in the context of other works, which I hope can be addressed prior to publication. These are outlined below.

Major comments
R1.1. It seems that the broader objective of this work is to reduce uncertainty in our understanding of global SOA budgets. If the authors wish to do the best possible job of addressing the SOA budget conundrum, then they could also bring in observations from other sources (e.g., EMEP, WSOC, etc). Indeed, use of measurement from IMPROVE and from 14 C dataset prove critical to this work's analysis. So if the real goal is to constrain the SOA budget, then perhaps additional data could be used, and the title of the paper adjusted. Otherwise, if the goal is rather to further understand what AMS alone can constrain, then more explanation / motivation along these lines might be beneficial.

A full analysis using all available OA, OC, WSOC observations would be an enormous task. The first step towards this that we have completed using the AMS and IMPROVE observations is already a substantial project, and one that has not previously been attempted. Including additional OA, OC, WSOC observations should be the next step but we strongly feel that this is beyond the scope of the present study. As suggested we include additional discussion and motivation. We add to the conclusions (P5725, L5):

"To enable tighter constraints from analysis with the methods presented here, requires bringing together OA, OC, WSOC observations from other sources along with additional observations in locations...."

We modify the existing text on P5703 to:

“In this paper we use a global aerosol microphysics model and a global dataset of OA observations from the aerosol mass spectrometer (AMS) to produce a new top-down constraint on the total and source-specific global SOA budget. The different SOA sources are parameterized based on results from recent field measurement campaigns, and the optimum is found by adjusting the strength of the different sources until the model/measurement error is minimized. We use OA observations from the IMPROVE network over the United States to test our optimised sources and a dataset of ^{14}C observations to evaluate our source-specific SOA budget.”

We prefer to stick with our current title as it is the most accurate description of our work, that is using the AMS data as the primary observational constraint on SOA budgets. Two advantages of the AMS are the consistency of using the same instrument at all the different locations, and that it provides information on OA components including OOA which is a closer link to SOA than bulk OA or OC. We add to P5707, L4:

“The identification of different OA components is useful for our analysis of SOA sources as we discuss in section 3.2”

We already showed that the SOA sources that were optimised against AMS data also appear to be consistent with IMPROVE data, which are completely independent. We agree that future optimization with more datasets is of interest and an important future step.

R1.2. The assumption that SOA forms direct and irreversibly from the all precursors could use more consideration. Would this assumption give an upper bound for local sources and an underestimate of any slower, longer scale formation processes? How would such a model account for observations of SOA aloft (e.g., Heald 2006), or for biogenically influenced background concentrations observed at the surface (e.g., Hodzic 2010)? It it perhaps by construction that the source attribution ascribes a bulk of the SOA to the sector with the largest overall influence in the northern hemisphere? So I wonder if the impacts of secondary (or higher) oxidation pathways, fragmentation and volatility could be brought more into the discussion.

The assumption that SOA is formed with a timescale of the order of a day or more is derived from the ambient observations. This timescale seems to match the ambient AMS and WSOC observations in multiple field campaigns. One can speculate about many possible mechanisms and processes, but we have chosen to implement a simplified model which in particular is consistent with the observations using fast instrumentation.

Longer-scale (in both time and space) formation processes that may contribute to a low concentration background are unlikely to explain the enhanced SOA concentrations that are consistently observed in polluted air as they will contribute mostly to a “smeared” background concentration. However, the referee is correct that for this reason these slower scale formation processes will be difficult to constrain with our method. To address these comments we have added the text to P5704, L24:

“We assumed these reactions form a surrogate oxidation product that can condense into the particle phase. The reaction rate of VOC_A with OH is based on field

observations (de Gouw et al., 2008; DeCarlo et al., 2010) and results in SOA production on timescales of ~1 day or more. Slower SOA formation processes (e.g., those from second or higher oxidation pathways) which will not lead to distinct spatial patterns in OOA concentrations will be more difficult to constrain with this method."

We already discussed observations that suggest that ambient SOA is much less volatile than that produced in the lab (ACPD P5705, L5-10). We already acknowledged (ACPD P5705, L13-14) that volatility is an additional parameter that needs to be further explored in this type of analysis. We modify/add text to P5705, L13:

"Although fresh SOA is known to be semivolatile (Odum et al., 1996; Cappa and Jimenez, 2010), it has been recently shown that aged SOA has low volatility (Jimenez et al., 2009; Cappa and Jimenez, 2010) and may form a highly viscous glassy state (Virtanen et al., 2010; Cappa and Wilson, 2010; Vaden et al., 2011), which is likely to be most relevant to the time and length scales of our global modelling study. Recently, Riipinen et al. (2011) examined particle formation events to show that roughly half of freshly-formed condensing organic vapours appears to be effectively non-volatile."

Observations of SOA aloft can be explained with the current paradigm, either by lofting of the SOA precursor or the SOA after formation. So we do not think that such observations are inconsistent with our model. However, there is limited evidence of substantial concentrations of SOA in the free troposphere (FT) and more recent papers (e.g., Heald et al. (2006), Dunlea et al. (2009), already cited in the ACPD version) suggest that underprediction of SOA in the FT may not be a consistent model feature. The clearest evidence for substantial SOA presence in the FT is the detection of isoprene organosulfates by Froyd et al (2010), but still these represent a minor fraction of the OA mass in FT.

We have already discussed issues surrounding OA in the FT. We have limited FT data, and SOA formed there is unlikely to make a major contribution to BL sites, so we do not think that is a major limitation of this analysis. We already estimated a substantial uncertainty in our optimised sources due to the limited observational data in the FT. Repeating this analysis using vertical profiles of OA and WSOC will be an important next step in this analysis.

R1.3. One of the main conclusions of the paper is that there is a large source of SOA that is biogenic in nature yet anthropogenically driven. And yet, the model simulations (24-32) including such properties did little to improve the model fit to the observations. This at first seems like a bit of a contradiction, despite the authors stressing that these simple schemes are not mechanistically representing SOA formation. Wouldn't it also be fair to say that none of the simulations, 1-32, mechanistically represent SOA formation, as all schemes considered are simplified empirical descriptions? Yet we are asked to believe that such simplifications are permissible for describing the SOA formation in simulations 1 - 23, so I found this a bit inconsistent. Are the authors suggesting that the anthropogenic mechanism for controlling the SOA is one that is entirely different than those presently considered, and fundamentally more complicated to represent than the direct formation pathways?

We agree - none of the simulations mechanistically represent SOA formation. It was not the intention of our study to elucidate a detailed mechanism of SOA formation, and this is likely not possible using a global model. Rather, we intended to constrain the global SOA source and if possible divide this source into several broad categories based on the spatial patterns of emissions and constraints on the correlations and timescales from recent field observations.

Our analysis appears to confirm previous suggestions that there is a link between SOA concentrations and anthropogenic pollution. We evaluated simplified versions of several different mechanisms (e.g., elevated SOA production through specified oxidation pathways, acid catalysed biogenic SOA formation) that have been suggested in the literature to explain such a hypothesis. These more explicit schemes were not more successful at explaining the ambient observations than a very simple scheme. We do not argue that the success of the simple scheme discounts any of these alternative mechanisms. However, in the absence of greater explanatory power we choose to use the simpler scheme. The hypothesized interactions in the more complex schemes may be what is going on, however the rates, yields, or functional dependences may not be well captured by our simple parameterizations. There are also maybe additional mechanisms that are possible that we did not try. The possible functional and parameter space is enormous, and it needs to be explored further with guidance from laboratory and field experiments. We modify the paper (see response to referee comment R1.6(a))

R1.4. While the authors clearly strive to place their results in context with prior works, I feel that the sentiment of previous modeling studies investigating anthropogenic SOA is not always well represented. I don't think there are any (or at least not many) modeling studies of SOA post 2005 which unequivocally claim that anthropogenic SOA is negligible relative to biogenic SOA. As chamber studies and field measurements have lead to new insights, modeling work has checked individual processes for their contribution to SOA on global scales. And yet, even when a single process has turned out not to be the smoking gun, the conclusions of these works is not that anthropogenic SOA is not important, just that the current process being tested by itself may not explain the larger budget issue. For example, Henze 2008 examines the amount of SOA produced from aromatic compounds via a specific mechanism, and these estimates shouldn't be misconstrued as representing a comprehensive model study of anthropogenic aerosol. Indeed, the conclusions of this work were similar to those of the present manuscript, as it notes "Consideration of additional mechanisms for formation of SOA beyond those considered here . . . would appear vital to our description of anthropogenic SOA." Similarly, the work of Farina 2010 notes that their findings "challenges the assumption that anthropogenic volatile organic compounds are not important SOA precursors on a global scale". Both works posit a total anthropogenic source possibly greater than 10 Tg / yr, so I think the wording on 5718, line 14, and again on page 5700, 22 could be adjusted. The present work moves beyond these by providing additional constraints.

It should also be noted that the perception in the community is still that biogenic SOA overwhelmingly dominates globally. However, we have revised the wording to more accurately describe recent reports of the few recent model studies of anthropogenic SOA formation. We reword the following sentence from the abstract "The urban/industrial SOA source is consistent with the 13 Tg a^{-1} estimated by de Gouw and Jimenez (2009), which was much larger than estimates from previous studies." to:

"The urban/industrial (fossil) SOA source is not inconsistent with recent estimates of the anthropogenic SOA formation."

We also change P5718, L14 to:

"The origin of the large anthropogenically controlled SOA source (100 Tg (SOA) a^{-1}, 70% of total SOA source) is unclear. This source is substantially more than that predicted by previous global model studies (e.g., Tsigaridis et al., 2006; Heald et al., 2008; Henze et al., 2008, Farina et al., 2010) and is not consistent with current understanding of SOA formation from anthropogenic VOCs."

We also add to the conclusions (P5724, L1):

"This anthropogenically controlled SOA source is greater than that in current global atmospheric models."

R1.5. 5711, line 17: How do you really feel about the fact that the best fit model was one in which you considered only remote sites and had SOA only from anthropogenic sources? Because we know this is a completely non-physical result, right? To me, this indicates that the inverse modeling framework is, by construction, not sufficiently sophisticated enough to capture the observations to which it is being compared, and this casts serious doubts on the rest of the analysis. Given the small amount of data, this is likely an ill-posed modeling task, which could amplify errors in the model or observations, and project these onto the solution (the source attribution). I suspect there is non-negligible error coming from representational error (see point 7) and the simplified SOA scheme (point 2), though possibly sampling plays a role (points 6a,b).

This is not an unphysical result since a) most remote sites, especially those in the Northern Hemisphere, are still influenced by pollution, e.g. they have elevated sulfate concentrations (so they may also have anthropogenically-controlled SOA advected or formed there, and in fact SOA and sulfate typically correlate over regional scales, see e.g. Zhang et al. (2005)); b) when we do not include biogenic SOA sources we do underpredict SOA at very remote sites such as AMAZE suggesting that most SOA at these sites is biogenic, c) we are suggesting that the anthropogenically controlled source is substantially larger than the biogenic source and dominates SOA even in many remote Northern Hemisphere sites, which are almost never isolated from the influence of pollution.

When we weight the observations in an attempt to correct for under representation we find an increased biogenic SOA source that improves the simulated SOA at very remote sites such as AMAZE.

We agree that the data is limited, but we have gone to considerable lengths to check that the results about source attribution are robust, including optimising the SOA sources using various different assumptions and checking the optimised sources against independent observations from IMPROVE. We agree that our analysis needs to be re-visited in the future with additional data, global and regional models, and by other groups. We add to the conclusions (P5734, L1) the following line to acknowledge that our source-resolved budget remains uncertain:

"The source-resolved budget is likely to have greater uncertainties than our estimation of the overall source. In particular, the biogenic and biomass burning sources are less well constrained because of limited observations in air masses that are strongly impacted by these sources."

We are aware of other efforts like Jathar et al., ACPD, 2011 but did not cite as the paper was published in APCD six weeks after our paper was submitted. We now include a citation to this paper (P5718, L20):

"Recent studies predict that primary anthropogenic emissions of S/IVOCs can produce substantial quantities of SOA (Pye and Seinfeld, 2010, Jathar et al., 2011)."

R1.6. linear model: I have a few questions and concerns about the linear response model. (a) 5712, 8: Simulations 18-33 were not included as they "did not result in improved model performance." I don't follow the reasoning here. Analysis of the model output vs observations is an entirely separate step from model reduction, as the goal with developing a response model is solely to represent a complex model with a simplified model, irrespective of any observations. Thus, to develop the best linear model, the largest space of input parameters should be explored, and weighted appropriately. Any additional information about how simulated SOA responds to changes in emissions, regardless of whether or not these SOA levels fit observations, helps the linear model characterize the global model. Regardless, there are many of the simulations in 18-33 that perform better than those in 1-17, so the distinction seems capricious. Rather, the decision about which simulations to include should depend upon the parameter uncertainty, not the model skill, as described in the next comment.

Those simulations were exploratory, trying to find the functional form of the model that best fit the observations. This was a screening step where we tried multiple functional forms based on previous literature. If we were to include them all in the linearized model, we would end up with too many degrees of freedom (i.e. sources) to fit. We chose to use the most simple form of the model as this simple form better matched the observations than more complicated forms. To better explain out method we add the following text to P5703, L11:

"In this paper we use a global aerosol microphysics model and a global dataset of OA observations from the aerosol mass spectrometer (AMS) to produce a new top-down constraint on the total and source-specific global SOA budget. We simulated OA using a 3-D global aerosol chemical transport model (section 2.1) and tested the model response to a range of different SOA sources (section 2.2). We used AMS observations (section 2.3) to evaluate these model simulations (section 3.1). This screening step was used to select the SOA sources that we used in our optimisation method (section 3.2). The optimization method involved building and running a simple linear model that reproduced the SOA simulated by the global model as a function of the different SOA sources. We then included the optimised SOA sources in the global model (section 3.5) and evaluated against both AMS observations and observations of

OA from IMPROVE. Finally we evaluated the optimized emissions using 14C radiocarbon observations (section 3.6).”

We also include the following text P5712, L7:

“We used as input the global model simulations 1-16 (Table 2) as these resulted in the best match with the AMS observations. We did not use simulations 17-33, where we explored potential pathways for anthropogenic enhancement of SOA formation from BVOCs, since they did not match the AMS observations as well as a simple model of anthropogenic SOA formation. Including all the formation pathways tested here would result in too many sources to fit given the limited availablility of AMS data to constrain the solution. We do not think that this analysis discounts the mechanisms tested in simulations 17-33, however the rates, yields, or functional dependences may not be well captured by our simple parameterizations. There are also likely to be additional mechanisms that are possible that we did not try. This is a large functional and parameter space, that needs to be explored further with guidance from laboratory and field experiments.”

R1.6(b) While there is ample literature describing how to build representative response models by sampling the parameter space according to the uncertainty and range of the parameters themselves, the sampling method employed here seems a bit ad hoc. The range of model states sampled from also seems limited and not equal across the parameters. The weighting of these samples according to the degree to which we believe the parameter values are reasonable is also not accounted for.

In response to the referee comments we have included additional analysis evaluating the response model against the global model. This analysis shows that our method is robust. Our method is applicable mainly because our global model is very linear with respect to changing emissions and our new analysis demonstrates this. If the model was less linear with respect to emissions more global model simulations sampling more of the parameter space would be required. We include the following discussion to P5712, L23:

“To evaluate the linear model we ran 6 additional global model simulations. We compared results from these simulations (which were not used to build the linear model) with output from the linear model. We found that the linear model well matches the global model (NME=0.6%, NMB=-0.17%, r^2=0.99). Since the system is linear and well behaved it is appropriate to use the simple linear model to emulate the behaviour of the global model.”

We realise that there are more sophisticated techniques to estimate the parameter values to use in the array of global model simulations and that this allows you to minimise the number of global simulations that are required to train the response model. However, in our technique we have fully sampled the parameter space (see Table 2). We build the linear model with 16 global model runs covering 5 parameters. We have demonstrated that our choice of approach is robust and will not impact our conclusions.

R1.6(c) No evaluation of the response model vs the full model is provided. This is a fundamental omission that needs to be provided if we are to evaluate any of the findings based upon the linear response model.

We were remiss not to include this evaluation and we thank the referee for spotting this. We now compare in the linearized and global model. We have run 6 additional global model cases that are not used as input for the linearized model. We compare results from these global model simulations with output from the linearized model. We add text to section 3.2, P5712, Line 23 (see response to R1.6(b)).

In section 3.5 we already test the emissions that were optimised in the linear/response model within the global model and compare against AMS observations.

R1.6(d) SOA formation is itself a nonlinear process. Can the authors estimate the degree to which the linear assumption is valid with respect to (a) the GLOMAP simulations and (b) what we know about more detailed SOA models?

In our model SOA formation should be linear with emissions. We now show the close level of agreement between the linear and global model demonstrating that the behaviour of the global model is very linear with respect to emissions.

Traditional models are indeed non-linear with respect to emissions, mainly because of the partitioning effects. However it is less clear that real-world SOA is as non-linear as the traditional models, as real-world SOA appears to show much lower volatility than the laboratory SOA in which the traditional models are based (e.g., Vaden et al., 2011, Cappa and Jimenez 2010, Riipinen et al., 2011). Our simulations are based on the field observations.

R1.7. There is considerable potential for representational error when comparing coarse global model estimates to a small number of surface observations, and it is not clear that this is handled in the best way. Throughout this work, model estimates are compared to the 47 individual AMS observations, and I don't believe this is the best use of a CTM in this context. Rather, the AMS observations should be aggregated so as to provide an estimate of the average concentrations within a given model grid box. The degree to which this grid box is fully / evenly sampled by the AMS observations would then be included as a "representational error" term in subsequent analysis. For example, grid boxes containing multiple AMS measurements would be given more credence than those containing just a single observation.

In this analysis there are no CTM model grid boxes that have multiple AMS locations within the same grid box. Our grid boxes are 2.8^ox2.8^o (200-300 km depending on the latitude). We interpolate model output to the location of the AMS observations rather than just using the grid box value (p5708, line 5-7). This accounts for spatial gradients in simulated concentrations.

R1.8. Regardless of issues of representational error or how the linear model is built, the method in which the response models are then used to optimize the sources in order to match the observations is a bit odd, and I'm not sure if it is theoretically sound. A much more

standard, and possibly effective, way to find the source terms would be to construct a matrix A that contains the coefficients a - e in each column and has 47 rows, one set of coefficients for each site. If the vector of observations is d and the vector of source terms is s = [SM; SI; SA; SBB; SP], then the optimal values of s are obtained directly from
s = (ATA).1Ad
where here optimal means minimizing the sum of least squares. To solve for s that minimizes the NME, a similar equation could be derived, though it would be nonlinear and require an iterative solution. If you used a standard discrete inversion method such as this, then you could also answer other questions such as: what is the uncertainty in you source estimate (A.1 Cd(A .1) T) and to what degree can different source sectors be resolved given the inversion setup (A .1 A), where Cd is the error covariance matrix of the observations.

The new analysis that we have completed here demonstrates that the system is well behaved and so it is appropriate to use the simple linear model. We have shown that the error caused by the linear model approach is small and substantially smaller than the model-AMS observation discrepancy. For this reason we prefer to stick with our original methodology. The approach described by the referee is definitely the best one in more complex inversion problems where there is much greater interaction between terms.

R1.9. Fig 6: To be honest, I have a hard time believing the isoprene values. With so many observations in the NE US, Europe, and Asia, and so few in isoprene hotspots such as SE US or the Amazon, it is difficult to believe that the isoprene source distribution is being similarly represented by the AMS locations.

We agree that high isoprene concentrations are underrepresented in the AMS dataset (see Fig 6b). We reword to account for this (P5714, L25):

"We found that simulated VOC concentrations at the AMS locations were broadly representative of the global simulated distribution of isoprene and monoterpenes (except that the highest concentrations of both are underrepresented), but are skewed to high VOC_A and low VOC_{BB} concentrations."

As the referee points out this is likely to be due to the AMS rarely being in SE United States or tropics. However, it should also be noted that whilst the AMS locations are biased to regions of lower isoprene the AMS observations are more often in the NH summer which is the local maximum for isoprene emissions. So these two factors will counteract each other to some extent and may be leading to more similar distributions than might otherwise be expected.

R1.10. 5716, 15: "does not greatly impact our results". This choice gives either a biogenic source of 0 Tg / yr or of 13 Tg / yr. I would say this is a huge difference, not a negligible one. Same reaction to this comment repeated in the conclusions.

We agree that this description was unclear in our original paper. We have modified the text to improve this point on P5716, L15:

“However, as we have shown our choice of method does not greatly impact our global total estimated SOA source.”

and P5725, L5:

“...and find that this does not greatly alter our global total SOA source”.

We already acknowledge that the biogenic and biomass burning sources are less well constrained (see P5716, L8-12).

R1.11. 5709, 5710: I know it’s consistent with what you expect that, for example, adding biomass burning will improve the performance. However, on the basis of the statistical metrics alone, this is hard to see. An r 2 of 0.0 and an r 2 of 0.03 both imply the model is largely incapable of describing the variance in the observations, and one may not be significantly better than the other. So I would perhaps place less emphasis on the differences in the correlation coefficient for any values less than 0.1, for which one might overall say the model has little correlation with the observations.

We have modified our paper (P5709, L21; P5710, L8; P5711, L10) so that we do not suggest that small changes in correlation coefficient are significant.

R1.12. 5702: What about cloud processing? Is there a reason the role of this process is not mentioned here, or throughout?

We now discuss cloud processing as suggested (P5702, L10):

“In cloud formation of SOA may be an additional source of SOA (e.g., Carlton et al., 2006; Chen et al., 2007; Tan et al., 2010).”

Our method gives little information on the mechanism, and if SOA is formed by cloud processing, it may still retain the correlation with CO on the scales of the global model. Including additional simulations with surrogate cloud processing mechanisms is of interest for future work, but outside the scope of the (already very long) current paper.

Technical corrections
R1.13. 5702, 20: It seems odd to cite Kroll 2006 as a reference supporting enhanced SOA formation in polluted regions, as this work showed that high NOx levels were inhibiting SOA formation from isoprene in such regimes.

We thank the reviewer for catching this error. We have updated this reference to “**(Chan et al., 2010; Ng et al., 2007b, 2008)**”, who showed increased SOA yields from isoprene under high NO_2 conditions, from sesquiterpenes under high NO_x conditions, and for isoprene reacting with the NO_3 radical, respectively.

R1.14. 5718, 17: Perhaps some clarification is required here, as the toluene and benzene yields in Ng 2007 were much less than 80%, more like 30-40%, and even then, only under low NOx conditions.

We agree that such high yields from anthropogenic VOCs are unlikely and we already stated that very conclusion in our ACPD paper (see P5718, L10-12). For improved clarity, we reword P5718, L17 to:

"Our optimised SOA source would require an average SOA yield of 79% from anthropogenic VOCs. Whilst high SOA yields (in the order of 30-40%) have been observed in laboratory studies under certain conditions for species such as for toluene and benzene (Ng et al., 2007; Hildebrandt et al., 2009), this yield is substantially larger than expected for the majority of anthropogenic VOCs."

R1.15. 5721, 11: 2009, not 2010, for the de Gouw and Jimenez reference.

We thank the reviewer for catching this error, which we have now fixed.

R1.16. 5701,16 and 5702, 11: I know what you mean (models that used alphas and K's in Odum fashion), but for others the terms "traditional" and "early" are vague and used inconsistently. I suggest you refer to methods by the mechanisms which they represent, such as "absorptive partitioning".

We define "traditional models" as those developed prior to 2006 (see P5701, L16-17). That point in time correspond to the publication of four studies in different regions and by different groups that pointed out that existing models underpredicted SOA formation in polluted regions by about an order of magnitude (see Volkamer et al., GRL, 2006 and references therein). The field of SOA modeling underwent intense growth after that point and many new models have been proposed since. Therefore we feel that the use of this term is justified, it is unambiguous since a definition is given, and it has been used in the same sense in several other publications.

For consistency we have changed "early" to "**traditional**" on the second instance mentioned by the reviewer.

R1.17. 5704: Why not just use the EDGAR VOC emissions rather than scaling the IPCC CO emissions to match the total EDGAR amounts?

Either approach could have been used and we do not think that this would greatly modify our results. Fossil fuel emissions of CO and anthropogenic VOC emissions are spatially well correlated. Field observations have linked observed OOA concentrations to CO, providing a motivation for using CO. Also CO inventories may be better constrained due to the availability of satellite measurements and the simplicity of modeling a single slow-reacting species, neither of which is true for VOCs.

R1.18. 5705, 3: "to reduce the computational burden of our simulations". It seems like a simplified scheme was beneficial for the sake of interpretation more than computational expense.

We have removed "to reduce the computational burden of our simulations " as suggested.

R1.19. 5708, 16: "no skill at capturing the spatial pattern". Wouldn't the r 2 value indicate skill at capturing both spatial and temporal variability?

There is limited temporal information in the data (many AMS observations are in summer and few sites where a seasonal cycle has been measured), but referee is correct and we reword to:

"no skill at capturing the spatial or temporal pattern".

R1.20. 5708, 16: "most likely because the spatial resolution of the global model is too coarse to resolve the urban-scale pollution." Can this be tested by repeating the analysis leaving out the urban sites?

We already did this when we restricted the analysis to remote sites (see P5709 and P5710 of ACPD version)**.**

R1.21. 5715 and Fig. 5: Are the biomass burning values in the figure off by a factor of 10?

No. Our method has trouble differentiating between biomass burning OOA and conversion primary POA to secondary OOA because both these sources are dominated by biomass burning and therefore have virtually the same spatial and temporal pattern. In Figure 5 the 2-D plane that we plot is for no POA to SOA source. With this assumption the biomass burning SOA source that best matches observations is greater. Increasing the POA to SOA source reduces the biomass burning source of SOA from VOC_{BB}. The optimum match when also including total OA as a constraint is a larger POA to SOA source and a smaller source from VOC_{BB}. We add the following to the paper P5717, L23:

"Our optimised SOA source from ageing of POA is greater than that directly from biomass burning VOC which is consistent with recent field observations (Cubison et al., 2011)."

Anonymous Referee #2

The manuscript describes a modeling study that aimed to reduce the uncertainty on the secondary organic aerosol (SOA) sources, both in magnitude and origin. The method is based on a multi-simulation approach, where several arbitrary selected cases were tested, and then were optimized, to match measurements. A different set of measurements was used to test the optimization results. The concept of this work is very good; it appears to be a very clever way to constraint SOA production. Nevertheless, there are some major issues that need to be addressed before this paper is accepted to ACP.

General comments

R2.1) The concept of this work is very good; the approach though seems not to be the proper one. The questions the authors tried to answer belong to an inverse modelling study, not a traditional one. Having measurements and trying to constraint sources is exactly what the inverse models are trying to constrain. Using a forward modeling technique like the one described in the manuscript, one trusts that the spatial distribution of sources is correct, and only the magnitude is the free parameter. This applies mainly to the primary sources of organic aerosols, but also at the precursor volatile organic compounds of SOA.

The referee is correct that we assume that the spatial pattern of the different precursor emissions is correct and that the major uncertainty lies in the conversion of these precursors to SOA. By considering a few sources of SOA (e.g., anthropogenic, biomass burning etc) we attempt to place some constraint on the source resolved budget. Our approach is unable to resolve the full spatial emission pattern of SOA precursors. An inverse modelling study that accounts for both the magnitude and spatial pattern of emissions is a complementary technique that would also be useful. A potential problem is that the limited observational dataset would not allow a good constraint on both the magnitude and spatial distribution of sources. We add the following test to further clarify our approach (P5705, L25):

"We assume that the spatial pattern of VOC and POA emissions are correct."

R2.2) A potentially significant source of SOA, that of IVOC, is not taken into account. In addition, the semi-volatile nature of primary OA is treated as a sensitivity study, and not as the default one. These two can significantly alter the OA spatial and temporal distribution, drastically affecting the results of this study. Especially concerning volatility, there are some very important free parameters in their calculation that need to be constrained, like the volatility distribution of the emissions, the temperature dependence of the volatility and their volatility change with time. Trying to estimate the sources of SOA using the traditional approaches appears to be problematic.

Anthropogenic IVOCs are implicitly included as part of our anthropogenic VOC source, since SOA formed from them will have a similar correlation with CO and timescale as for other anthropogenic organic gases (see e.g. Dzepina et al., ES&T, 2011, who show this result explicitly using the currently available parameterizations of SOA formation from IVOCs). Similarly for biomass burning IVOCs which are implicitly included in the VOC_{BB} source.

The comparison with the ^{14}C data helps constrain the contribution from fossil sources, including anthropogenic IVOCs, and suggests that it can not be much more than 10 Tg SOA a^{-1}. We already treat SOA formation from SVOCs implicitly through the VOC_A, VOC_{BB}, and aging of POA sources. We show the aging of POA to be an important global source, especially for biomass burning POA.

Regarding volatility, see responses to comments R1.2 and R1.6(d). We already state that volatility of SOA needs to be explored in future studies of this type. We modify our paper to emphasise that we assume that SOA is non-volatile. We add to the abstract (P5700, L9):

"We assume that SOA is essentially non-volatile and condenses irreversibly onto existing aerosol".

We already discuss this assumption and suggest that volatility needs to be explored in future studies of this type (see P5705, L3-14). We modify our paper by adding references to recent studies (e.g., Riipinen et al., 2011; Vaden et al., 2011; Cappa and Jimenez, 2010) that suggest that aged SOA is rather non-volatile.

R2.3) It is not very clear where the anthropogenically controlled SOA come from. The authors mention that they are linked to anthropogenic emissions, but only 10% is urban.

Where does the carbon come from? If only from anthropogenic sources, then the 100Tg/a of SOA seem excessive, even with an OM/OC ratio of 2. This would mean that almost all primary anthropogenic emissions have a close to 100% SOA yield. On the other hand, if the carbon comes from biogenic sources that mix with anthropogenic ozone, NOx and other pollutants, then this source double-counts the biogenic carbon fluxes, since they are also being considered as monoterpenes and isoprene. Scaling the IPCC CO emissions from which sources gave the VOCa emissions distribution?

We already discussed the SOA yield that would be required for anthropogenic VOCs if all the anthropogenically controlled SOA came from this source (see P5718, L15 of the ACPD version). We used the ^{14}C data along with the model (see section 3.6, P5719) to show that at most ~10% (10 Tg SOA a^{-1}) of the anthropogenic controlled SOA is from fossil sources (anthropogenic VOC and IVOCs). We have used the term “urban SOA” to describe this source. The remainder of the carbon from the anthropogenically controlled source must come from contemporary carbon sources.

We clarify our discussion of potential sources of anthropogenic SOA, and modify our description of anthropogenically controlled SOA (P5703, L25) to:

“Throughout the paper we term the SOA that is linked to anthropogenic activity as “anthropogically controlled” to underscore the fact that it could represent any source with a spatial pattern similar to that of fossil fuel CO emissions which may include the following a) SOA formed directly from oxidation of anthropogenic VOCs, b) IVOCs, c) enhancement of biogenic SOA production due to anthropogenic pollution, d) meat cooking, e) wood smoke emissions that are not in standard emission inventories. We use ^{14}C data in an attempt to differentiate contemporary from fossil sources as described below. However, our analysis method provides little mechanistic information that can help determine the dominant formation pathway.”

We also clarify (P5718, L10) to:

“The origin of the large anthropogenically controlled SOA source (100 Tg (SOA) a^{-1}, 70% of total SOA source) is unclear. This source is substantially more than that predicted by previous global model studies (e.g., Tsigaridis et al., 2006; Heald et al., 2008; Henze et al., 2008, Farina et al., 2010) and is not consistent with current understanding of SOA formation from anthropogenic VOCs. If this source was solely from anthropogenic VOCs this would require an average SOA yield of 79%.”

Option (c) does not double-count the carbon sources, as the results of the model are constrained by the observations. In this case there is a fraction of BVOC that appears to convert to SOA under ”background” conditions and an additional fraction that converts under polluted conditions. We calculate what the combined yield from biogenic carbon would be in each case.

We use IPCC CO emissions from fossil fuel sources to provide the spatial pattern of “anthropogenically controlled” SOA sources.

R2.3(b) How can it be verified that scaling CO in one hand and using aromatic emissions in the other hand does not double-count carbon sources?

This approach does not double-count carbon sources. We could have chosen to use both the spatial pattern and magnitude of emissions of anthropogenic VOC from EDGAR. Instead we chose to use the spatial pattern from fossil fuel CO emissions and the global magnitude from EDGAR VOCs. We did this firstly because of the strong and consistent observed relationship between CO and SOA in field studies (see e.g. de Gouw and Jimenez, 2009; DeCarlo et al., 2010, cited in the manuscript) and secondly because this source is meant to represent several potential SOA sources from anthropogenic activities. We already clearly stated our methods in the ACPD version.

R2.4) The whole study is very strongly related with the host model. Such fine tuning of sources depends on the underlying climate, transport, chemistry, removal. A critical question which was not addressed is how much this optimization depends on the model. Understanding that it would be unrealistic to repeat this experiment with another model, one should try to implement a few key simulations from table 2 to a very different global model, and compare the results with the present study. On top of that, it should be extremely clear in the manuscript (abstract, body, conclusions) that this optimization is not meant to be a generic suggestion; rather, it is a very model specific calculation of some, but not all, uncertainties in global SOA sources. If a second model is included in the study, then, depending on the results, one might extract a more general conclusion. At the present state of the manuscript, the authors should discourage the reader from using their numbers as a suggested range of SOA sources.

We feel that it is beyond the scope of this work to test our method in another model. Use of multiple models within one study is not the standard methodology used within our community except within specific model intercomparison studies such as AEROCOM. We agree that our study needs to be repeated by other groups in the community using different models and we are aware of 2 modelling groups that are proposing to test our optimised emissions in their models. As suggested we modify both the abstract, P5700, L3:

“and the GLOMAP global chemical transport model”

and conclusions, P5723, L14:

“We have used the GLOMAP 3-D chemical transport model”.

R2.5) The SOA production chosen, as shown in Table 1, does not agree with literature data. It had been shown by numerous publications the last 15 years that using a single product to simulate SOA formation from a given precursor VOC is insufficient.

Very complex models of SOA (including 1000’s of reactions) still do not well explain ambient SOA concentrations and such models are too computationally expensive to include in global CTMs or climate models. Our approach is to use a simple parameterised description of SOA based on the correlations, timescales, and volatility of the field observations, which are documented in the literature (see e.g. de Gouw and Jimenez, 2009; DeCarlo et al., 2010, cited in the manuscript). Our parameterization is easily tunable to best match ambient SOA observations. We find that ambient SOA concentrations are “reasonably” explained by our simple SOA production mechanism with a relatively short timescale (in the order of one day). We do not suggest that our parameterised mechanism matches the real mechanism that occurs in the atmosphere. However, we think that that the agreement between model and observations does suggest that the real mechanism must rapidly produce SOA on a timescale such that SOA production may occur at sub-grid (<300 km) scales.

R2.6) Given the large uncertainties of the study, I suggest to remove section 3.7. The SOA sources are so heavily tuned and fit to the GLOMAP model, that calculating a radiative forcing out of it appears premature.

We prefer to give the readers an order-of-magnitude impression for the potential climate impact of such a source. We already include an error estimate of the radiative forcing. We agree with the referee that there are substantial uncertainties and we modify our paper to emphasise these and suggest that the error may be larger than estimated. We remove the radiative forcing numbers from the abstract so that they are not separated from our discussion of the uncertainties. We add discussion to emphasise the uncertainties. We add text to P5721, L22:

“We estimate an order of magnitude radiative effect due to the presence of anthropogenically controlled organic aerosol. We estimate both the aerosol direct effect (ADE) and the cloud albedo (first) aerosol indirect effect (AIE) through …..”

We modify and include additional discussion on the uncertainties P5723, L8:

“There is considerable uncertainty in our estimate of the potential radiative effect due to anthropogenically controlled SOA, due to uncertainty in the mass budget and the optical and cloud nucleating properties of the aerosol. There is also considerable uncertainty due to the volatility distribution of the SOA which controls where in the aerosol size distribution the material will condense (Riipinen et al., 2011). We have assumed essentially non-volatile SOA which likely provides an upper limit for the contribution of the material to CCN concentrations and leads to a substantial and negative AIE. If the material was partially semi-volatile, condensation on larger particles would be enhanced resulting in reduction in nanoparticle growth rates and an enhancement of the condensational sink which would suppress nucleation. This would suppress and potentially reverse the enhancement to CCN and CDNC and hence lead to a smaller negative or even positive AIE. This can be seen in the study of Arneth et al. (2010) who calculated that SOA (largely from biogenic sources) contributes a global mean all-sky ADE of -0.08 W m^{-2} but a positive AIE of 0.3 Wm^{-2}. Future work is needed to explore uncertainties in the volatility and contribution of SOA to particle growth rates and the interaction with BL particle formation (Spracklen et al., 2008a) both of which may greatly impact the contribution to the formation of CCN and the AIE.”

We also add the following text to the conclusions: P5724, L22:

“We estimate the order of magnitude of the radiative effect due to our proposed anthropogenic SOA source. Substantial uncertainties exist in our estimate due to

uncertainty in the anthropogenic SOA budget, the optical and cloud nucleating properties of the anthropogenic SOA and the volatility distribution of the SOA which controls where in the aerosol size distribution the material will condense (Riipinen et al., 20011).”

and P5724, L27:

“Future studies using different models are now needed to confirm these estimates.”

Specific comments
R2.7) Pye and Seinfeld (2010) are not the only ones that made a global modeling study on IVOC and the semi-volatile organics, Jathar et al. (ACPD, 2011) have also done one.

Jathar et al. was published in APCD six weeks after our paper was submitted to this journal, and we did not have access to that paper before it was published, and this is the reason why it was not cited in our ACPD manuscript. We include a citation to this paper (P5718, L20):

“Recent studies predict that predict that primary anthropogenic emissions of S/IVOCs can produce substantial quantities of SOA (Pye and Seinfeld, 2010, Jathar et al., 2011).”

R2.8) Page 5704, lines 8-10: Do you have such a process in the model?

Yes, we include this process in a simple way (see P5704, L4-L6 of the ACPD version).

R2.9) Page 5704, lines 15-18: Is this scaling justified for non-polluted regions? How can a correlation of CO and aerosols be justified, given their very different lifetimes?

The scaling is based on emissions, and not on ambient concentrations, so the difference in lifetimes is not an issue.

R2.10) Page 5704, lines 18-22: The authors seem not to trust the EDGAR global numbers, but do use their spatial distributions. How can this be justified?

Either approach could be justified. We clearly explain our approach (see P5704, L14-L22 in the ACPD version). See also response to comment R1.17.

R2.11) Page 5705, first paragraph: Volatility is extremely important. Pye and Seinfeld (2010) and Jathar et al. (2011) get different answers by using the volatility change with aging differently. Tsigaridis and Kanakidou (2003) get different vertical distributions by changing just the temperature dependence of volatility, and about a factor of 3 increase in SOA production by not allowing them to evaporate. In a global modelling study, aerosols might have plenty of time to evaporate and recondense. Please comment on the uncertainty introduced by assuming no evaporation.

The referee is correct that volatility distribution is a substantial uncertainty and plays a major role in model studies. Future work needs to test whether more complex models can explain more of the variability in the observations than the simple mechanisms used here. There is little point adding additional complexity to the model if the new model introduces more uncertain model parameters and the observations from the ambient atmosphere can't constrain it. We feel that using the simplest model which can explain the observations is a very valuable exercise. We already acknowledge that volatility needs to be explored in future studies of this type (see P5704, L13-14). However, as we discuss (P5705, L3-L10) ambient studies suggest that atmospheric SOA is much less volatile than previously thought. The differences between the Pye and Jathar studies can be better described as differences in SOA yields of their mechanisms, rather than volatility (which is just an intermediate parameter). We modify and add to our discussion (see response to referee comment R1.2). We also add the following text to the conclsions (P5723, L25):

"In this study we assumed SOA had negligible vapour pressure and condenses irreversibly onto pre-exisiting aerosol. This is consistent with recent observations that suggest that aged SOA has low volatility (e.g., Jimenez et al., 2009; Cappa and Jimenez, 2010; Vaden et al., 2011; Riipinen et al., 2011). However, future studies need to explore the sensitivity of the calculated SOA budget due to uncertainties in volatility."

R2.12) Table 2 is very hard to read. It belongs to an appendix, and it has to be replaced by a more condensed and elegant presentation of the simulations performed and their results. In addition, the criteria used in the selection of simulations have to be mentioned. All possible combinations of cases have been simulated? If no, which were the criteria used for selecting which ones to simulate? How was it tested that no bias was introduced by not simulating every possible combination?

We feel that table 2 is important and needs to stay in the body of the paper. It is impossible to simulate all possible combinations of SOA sources, even in a simplified framework such as ours. We run 33 global model simulations. Simulations 1-16 are used to build the global model. Simulations 17-33 are used to test potential mechanisms for anthropogenic enhancement of SOA production from BVOCs. We selected the simulations 1-16 in order to sample the parameter space with respect to the 5 different SOA sources. To test that no bias was introduced by our choice of simulations we now evaluate the linear model against additional global model simulations (that were not used to build the linear model). See our response to referee comments R1.6(c) and R1.6(d).

R2.13) Page 5706, line 5: How many Tg/a of SOA from monoterpenes?

This simulation includes 32.3 Tg a^{-1} SOA (see Table 2, experiment 1).

R2.14) Page 5707, lines 1-4: Several papers presenting AMS results show a third component, in addition to HOA and OOA, that of BBOA, from biomass burning. Is it valid, based on measurements, to assume that 100% of BBOA is HOA?

For the purposes of the present study and for simplicity, BBOA and also other primary components have been lumped into HOA, which represents all POA. Similarly many AMS studies have identified several types of OOA (see e.g. Jimenez et al., 2009), but they are all lumped into a single OOA for the purposes of our study. We already state this in the ACPD version of the paper (P5706, L24 - P5707, L4).

R2.15) Page 5707, lines 13-15: How about Heald et al. 2005 that shows a very large discrepancy between measurements and models in the free tropospheric OA?

More recent studies (Heald et al., 2006; Dunlea et al., 2009) suggest that the very large discrepancy that was seen in Heald et al. (2005) may not be a widespread phenomenon. This was already discussed and these papers cited in the ACPD paper (see P5707, L13-L15). We already acknowledged that lack of information on the vertical profile of OOA is a limitation of our study (see P5707, L15-L16 and P5724, L29 - P5725, L1 of the ACPD version).

R2.16) Page 5708, lines 11-13: The model appears to be as "successful" with HOA as it is with SO4.

We modify the paper to reflect this.

R2.17) Page 5709, lines 17-23: The tropical AMS measurements (OP3 and AMAZE) were during the wet season, thus no biomass burning. Which other measurements were used to constrain the biomass burning source in the tropics? If none, then the global biomass burning source was constrained from mid-latitude and boreal fires, which is not correct.

There are observations from DABEX and AMMA field campaigns (see Table 3). We clarify this by modifying the paper (P5707, L20):

"Observations from the AMMA and DABEX fields campaigns sample air influenced by tropical biomass burning."

We already acknowledged that the biomass burning source is not well constrained (e.g., see ACPD abstract in P5700, L25-L27 and P5716, L7-L13) and that we need more measurements in air influenced by biomass burning (both in the tropics and extra-tropics) (e.g., see ACPD Page 5725, L5-L9).

R2.18) Page 5709, line 27: "heterogeneous oxidation": why not homogeneous oxidation as well?

We modify the paper to clarify that both heterogeneous and homogenous oxidation could be occurring (P5709, L27):

"This overprediction could be due to homogeneous and heterogeneous oxidation...."

R2.19) Page 5713, line 29 to page 5714, line1: I disagree with this statement. Most models have the same aging parameterization, which is that of Cooke et al.: Globally constant lifetime against aging of about a day. The real uncertainty should be much larger.

We modify paper to say that the real uncertainty may be even larger than that suggested by the model diversity (P5714, L1):

“although we acknowledge that the real uncertainty could be greater than that indicated by the model diversity.”

R2.20) Page 5714, line 3: The sentence should start with “We arbitrary estimated”. If the choice was not arbitrary, it should be justified.

We use the word “estimated” to inform the reader that our choice of error is not exact.

R2.21) Page 5715, line 8: “To remove” should be “To correct”. The bias cannot be eliminated, since lifetime affects the distribution.

We modified the text as suggested.

R2.22) Section 3.4: Is the optimization affecting the lifetime of OA? If yes, how much? If no, why?

The change in lifetime is small as we do not modify the removal mechanisms or the spatial pattern of emissions.

R2.23) Page 5716, lines 15-16: For SOAb, there is a factor of 5 difference. It might not be so important globally, but it clearly is regionally.

We modified the text as suggested; see response to referee R1.10 for details.

R2.24) Page 5717, last paragraph: Not all authors use the same definitions and precursors for SOA. Please confirm that the comparison is valid and mention it in the text.

We believe that the paragraph is clear and that the definitions are consistent with those of the cited papers.

R2.25) Page 5719, 2nd paragraph: How do the 14C measurements correlate with IMPROVE and AMS? Is there any overlap between them?

There is little overlap between 14C and the other measurements as historically the measurements were performed separately by different groups, although that is starting to change as more field campaigns start to incorporate both AMS and 14C measurements. Aiken et al. (ACP 2010) present an intercomparison between the AMS OC and the OC measurements from two sets of filters, including those used to measure 14C, showing reasonable agreement.

R2.26) Page 5720, lines 17-19: The 10% urban SOA appears to be arbitrary. See also general comment #3.

We respectfully disagree. This is not arbitrary at all, but it is deduced from the comparison with the ^{14}C data (see section 3.6 in the ACPD manuscript).

R2.27) Page 5721, line 2: “substantially larger enhancement”: how much?

Almost an order-of-magnitude (see P5721, L5-L7 in the ACPD manuscript).

R2.28) Table 1, reactions 9 and 10: Are these the ones that are used to simulate the anthropogenically influenced SOA? Why regenerating VOCa and SO2? Why there is no oxidant dependence? How important these reactions are to the global SOA? More explanations are needed here.

No. These are sensitivity studies that are not used in the optimisation. Reaction 7 is used to create anthropogenically controlled SOA. We modify our paper to clarify this on P5706, L9:

"In the fourth set of experiments (simulations 11–24, reaction 7) we included SOA from anthropogenic VOCs..."

R2.29) Figure 8a,d: There is a peak above Mexico. Is this due to the Mexico City very high measurements? Maybe it biases the whole optimization high?

The referee is correct that the peak in simulated concentrations is due to emissions in and around Mexico City. However this does not bias our optimisation high since we have demonstrated that simulation of both the polluted/urban and the remote observations are improved by using our optimised emissions, of which Mexico City is only one 47 datapoints (e.g., see P5711, L17-L21 in the ACPD version).

R2.30) Figure 9d: Aged POA can be seen downwind of biomass burning regions, but not downwind of anthropogenically influenced regions. Is this a color scaling issue?

Yes, this is a limitation of the colorscale, due to the much larger emission of POA from biomass burning compared to fossil fuel sources.

Technical corrections
R2.31) Section 3.3 should be before 3.2, there is no point optimizing the non-representative results.

We prefer to keep our current approach in the manuscript. We optimise the emissions against the observations as they are. We then attempt to remove bias in the location of the AMS observations and re-optimise the source. If we only optimised against the weighted observations it would be unclear whether the results were an artefact of our weighting procedure. By optimising against both the un-weighted and weighted observations we demonstrate that our main results are robust, and what the main impact of the non-representativeness of the initial approach was (lower biogenic SOA but little difference in the other sources).

R2.32) Page 5713, lines 8 and 19: "OOA" should be "SOA".

Strictly speaking POA oxidation produces SOA if the oxidation happens in the gas-phase (from evaporated POA species) and produces oxidized POA (OPOA, also part of OOA) if the oxidation happens in the particle phase (heterogeneously). As the former path is thought do dominate (e.g. Robinson et al., Science, 2007) and for simplicity for the current manuscript, we modify the text as suggested by the reviewer.

R2.33) Page 5714, line 15: "may introduce" should be "introduces".

We modify as suggested.

R2.34) Page 5717, line 7: "United States" should be "western US in summer".

We modify as suggested.

R2.35) Table 2 legend: " SOA yields: : :percentage": I do not understand this sentence.

We have changed the text for clarity to:

"SOA yields (y1–y10) are defined as the SOA mass formed as a fraction of the emitted VOC mass expressed as a percentage"

R2.36) Figure 6: The concentration bins are extremely large.

This is necessary given the number of data points available.

Additional references for these responses:

Arneth, A., Harrison, S. P., Zaehle, S., Tsigaridis, K., Menon, S., Bartlein, P. J., Feichter, J., Korhola, A., Kulmala, M., O'Donnell, D., Schurgers, G., Sorvari, S., Vesala, T.: Terrestrial biogeochemical feedbacks in the climate system, Nature Geoscience, 3, 525-532, 2010.

Cubison, M. J., Ortega, A. M., Hayes, P. L., Farmer, D. K., Day, D., Lechner, M. J., Brune, W. H., Apel, E., Diskin, G. S., Fisher, J. A., Fuelberg, H. E., Hecobian, A., Knapp, D. J., Mikoviny, T., Riemer, D., Sachse, G. W., Sessions, W., Weber, R. J., Weinheimer, A. J., Wisthaler, A., and Jimenez, J. L.: Effects of aging on organic aerosol from open biomass burning smoke in aircraft and lab studies, Atmos. Chem. Phys. Discuss., 11, 12103-12140, doi:10.5194/acpd-11-12103-2011, 2011.

Froyd, K.D., Murphy, S. M., Murphy, D. M., de Gouw, J. A., Eddingsaas, N. C., Wennberg, P.O., Contribution of isoprene-derived organosulfates to free tropospheric aerosol mass, P. Natl. Acad. Sci. USA, 107, 50, 21360-21365, 2010.

Jathar, S. H., Farina, S. C., Robinson, A. L., and Adams, P. J.: The influence of semi-volatile and reactive primary emissions on the abundance and properties of global organic aerosol, Atmos. Chem. Phys. Discuss., 11, 5493-5540, doi:10.5194/acpd-11-5493-2011, 2011.
Ng, N. L., Kroll, J. H., Chan, A. W. H., Chhabra, P. S., Flagan, R. C., and Seinfeld, J. H.: Secondary organic aerosol formation from m-xylene, toluene, and benzene, Atmos. Chem. Phys., 7, 3909--3922, \href{http://dx.doi.org/10.5194/acp-7-3909-2007}{doi:10.5194/acp-7-3909-2007}, 2007a.

Ng, N. L., Chhabra, P. S., Chan, A. W. H., Surratt, J. D., Kroll, J. H., Kwan, A. J., McCabe, D. C., Wennberg, P. O., Sorooshian, A., Murphy, S. M., Dalleska, N. F., Flagan, R. C., and

Seinfeld, J. H.: Effect of NOx level on secondary organic aerosol (SOA) formation from the photooxidation of terpenes, Atmos. Chem. Phys., 7, 5159-5174, doi:10.5194/acp-7-5159-2007, 2007b.

Riipinen, I., Pierce, J. R., Yli-Juuti, T., Nieminen, T., Häkkinen, S., Ehn, M., Junninen, H., Lehtipalo, K., Petäjä, T., Slowik, J., Chang, R., Shantz, N. C., Abbatt, J., Leaitch, W. R., Kerminen, V.-M., Worsnop, D. R., Pandis, S. N., Donahue, N. M., and Kulmala, M.: Organic condensation: a vital link connecting aerosol formation to cloud condensation nuclei (CCN) concentrations, Atmos. Chem. Phys., 11, 3865-3878, doi:10.5194/acp-11-3865-2011, 2011.

Tan, Y., Carlton, A. G., Seitzinger, S.P., Turpin, B. J., SOA from methlyglyoxal in clouds and wet aerosols: Measurement and prediction of key products, Atmos. Env., 44 (39), 5128-5226, 2010.

Vaden, T. D., Imre, D., Beranek, J., Shrivastava, M., Zelenyuk, A.: Evaporation kinetics and phase of laboratory and ambient secondary organic aerosol, PNAS, 108, 6, 2190-2195, 2011.

Zhang, Q., D.R. Worsnop, M.R. Canagaratna, and J.L. Jimenez. Hydrocarbon-like and Oxygenated Organic Aerosols in Pittsburgh: Insights into Sources and Processes of Organic Aerosols. *Atmospheric Chemistry and Physics*, 5, 3289-3311, 2005.